

AD \_\_\_\_\_

GRANT NO: DAMD17-94-J-4420

TITLE: Crystallization and Structure Determination of the Human Estrogen Receptor by X-ray Diffraction

PRINCIPAL INVESTIGATOR(S): Stephen C. Harrison, Ph.D.

CONTRACTING ORGANIZATION: Children's Hospital  
Boston, Massachusetts 02115

REPORT DATE: October 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;  
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

19960315 099

DTIC QUALITY INSPECTED 1

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 1995	3. REPORT TYPE AND DATES COVERED Annual 15 Sep 94 - 14 Sep '95	
4. TITLE AND SUBTITLE Crystallization and Structure Determination of the Human Estrogen Receptor by X-ray Diffraction			5. FUNDING NUMBERS DAMD17-94-J-4420	
6. AUTHOR(S) Stephen C. Harrison, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Childrens Hospital Boston, Massachusetts 02115			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  The project is a postdoctoral training fellowship for Dr. Robert Nolte, to support X-ray crystallographic studies of cancer-related macromolecular complexes. (1) We have expressed various constructs of the human estrogen receptor DNA-binding domain, but failed to obtain crystallizable protein. Success of other groups in this and related efforts have led us to focus on different goals. (2) We have prepared crystals of a complex between a six Zn-finger segment of the transcription factor TFIIIA and a 31 bp DNA site. This structure will reveal significant aspects of DNA recognition by Zn-finger proteins, which regulate many loci important for cell growth. (3) During the coming project year, Dr. Nolte will embark on structural studies of Src-related tyrosine kinase activity, as described in the body of the progress report.				
14. SUBJECT TERMS Estrogen Receptor, Diffraction, Steroid, Hormone, Breast Cancer			15. NUMBER OF PAGES 18	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

X Where copyrighted material is quoted, permission has been obtained to use such material.

X Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

V Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.


       In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

       For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

V In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

X In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

 12 Oct 95  
PF - Signature Date

## **TABLE OF CONTENTS**

<b>A. Introduction</b>	2
<b>B. Body of Narrative:</b>	5
1. Estrogen Receptor	
2. TFIIA	
3. Src-family kinases	
<b>C. Conclusion</b>	9
<b>D. References</b>	10

## **Abstract**

The project is a postdoctoral training fellowship for Dr. Robert Nolte, to support X-ray crystallographic studies of cancer-related macromolecular complexes. (1) We have expressed various constructs of the human estrogen receptor DNA-binding domain, but we have failed to obtain crystallizable protein. Success of other groups in this and related efforts have led us to propose a modified statement of work, as outlined below, consistent with the central training goals of this project. (2) We have prepared crystals of a complex between a six Zn-finger segment of the transcription factor TFIID and a 31 base pair DNA site. This structure will reveal significant aspects of DNA recognition by Zn-finger proteins, which regulate many loci important for cell growth. (3) During the coming project year, Dr. Nolte will also embark on structural studies of Src-related tyrosine kinase activity, as described in the body of the progress report.

## **Introduction**

The goal of this project is postdoctoral training for Dr. Robert Nolte in the area of structural biology and more specifically X-ray crystallography as applied to problems in cancer and to breast cancer in particular. It was proposed to crystallize the hormone binding domain (HBD) of the estrogen receptor, in order to determine its structure and the implications of structure for the mode of hormone recognition. As summarized more fully in the body of this report, a number of constructs were prepared, without success in obtaining soluble crystallizable protein. Because other groups have now achieved similar goals, (Bourquet et al., 1995; Renand et al., 1995; Wagner et al., 1995), we request a change in the statement of work (SOW) to two other, related projects in the structural biology of cancer, in order to ensure rapid progress in Dr. Nolte's training, especially in X-ray crystallographic structure determination. The proposed new SOW and its relation to the old SOW are outlined below.

The problem : Cancer involves anomalies in cellular regulation. Two key regulatory steps are the transduction of signals from the cell surface and the control of specific gene expression. The estrogen receptor, on which the original SOW focused, participates in both these steps. It receives hormonal signals and in response, activates defined genes. The recently discovered BCRA1 gene appears to encode a nuclear protein that may regulate gene expression (Miki et al., 1994). Progress in structural studies of tyrosine-kinase mediated signaling pathways and protein/DNA complexes that regulate gene expression makes the possibility of structure-based drug discovery and development a real one, but many basic principles remain to be discovered. The modified SOW outlined here focuses on different cases of gene regulation and signal transduction. The goal remains to use X-ray crystallography to uncover the basic molecular structural features governing these processes and to train Dr. Nolte in this field.

## Previous work

(i) A number of structural studies of protein/DNA complexes involved in transcriptional regulation have provided the guidelines of how to think about specific recognition (see, for example, Harrison, 1991 and Steitz, 1990). An important notion to emerge both from the structural studies and from studies of transcriptional regulation at many complex promoter/enhancer sites is that large multi-protein complexes bound to DNA are critical for the sort of combinatorial control seen in cells of all higher eukaryotes (and in human cells in particular). Work in our laboratory on the polymerase III transcription factor known as TFIIIA has led to crystallization of a complex between a part of the factor containing six "zinc fingers" (each finger is a distinct DNA-binding domain) and a 31 base pair DNA site. This is a larger complex than any so far studied from this class of transcriptional regulatory proteins.

(ii) Protein tyrosine kinases are responsible for many key signaling events in normal and cancer cells. The kinases related to Src (so designated because of its discovery in Raus sarcoma virus) have been especially widely studied, in part because of the early importance of Src in revealing the existence of oncogenes and oncogene products. We have determined the structure of the Src-homology 2 (SH2) and Src-homology 3 (SH3) domains of lymphocyte kinase (Lck), and we are engaged in an effort to crystallize the auto-inhibited form of the full Lck protein. An important adjunct to this effort, which could spawn a number of drug-design strategies, will be to apply the expression approaches developed for Lck to another Src-like kinase such as Fyn, as well as to the protein that is "next" in the Lck-initiated pathway of T-cell signaling, zeta associated protein 70 (ZAP-70).

### **Purpose of present work**

The broad goal is to enhance our understanding of specific protein-protein and protein-DNA interactions in gene regulation and in signal transduction. In particular, we are making rapid progress in the X-ray crystallographic structure determination of a complex between a larger DNA-binding segment from the polymerase III regulatory factor TFIID and its cognate site. We are also building upon our own recent success in structural studies of regulatory domains from Src-like protein kinases to establish a broader program. Thus, Dr. Nolte will first complete the TFIID/DNA structure determination (in collaboration with Dr. Ray Brown of this laboratory), a task that should take about 9 months. He will then join our effort on Lck and initiate a parallel project on either Fyn or ZAP-70, as the state of our overall project requires. The cancer relevance of these studies is described in the body of the report.

### **Methods.**

All of the projects described here involve expression of proteins, or of specific protein domains, in order to crystallize them for investigation by X-ray diffraction. X-ray methods are the best way to determine the detailed atomic structure of a protein, a necessary pre-requisite for molecular analysis of function and drug-targeting strategies.



## **Body of Narrative**

**Estrogen Receptor:** The estrogen receptor (ER) plays a significant role in the onset and treatment of breast cancer. Details of its structure could potentially lead to the design of new therapeutic agents and increase our understanding of the role this receptor plays in the onset of cancer. The goal of the first year of this training grant was to express and crystallize the hormone binding domain of the ER and use of X-ray crystallography to determine the molecular details of the binding of estrogen and anti-estrogen compounds to the ER.

Previous work in this laboratory had concentrated on constructs containing only the hormone-binding domain (HBD) beginning at either residue 297 or 342 and ending with residue 551 or 590 (Figure 1). These four constructs failed to yield soluble protein, as required for the proposed crystallographic studies. During the first year of this grant protein expression constructs were made containing larger portions of the ER in an attempt to refine the domain boundaries of the HBD under the premise that the previous domain boundaries did not encompass the complete hormone binding domain. The constructs were made using a polymerase-chain reaction (PCR) method to obtain the DNA sequences corresponding to the fragments of interest. The fragments were subcloned into a pRSETa expression vector (Invitrogen), using the Nhe1 and Bamh1 restriction sites, and then transfected into *E. coli* strain BL21. A summary of the expression constructs created during the first year of this grant period is shown in Figure 1. The constructs are labeled to represent the different modular domains that have been identified in the nuclear receptor family. Little is known about A/B,D, and F domains. The C domain is the DNA binding module, and the E domain is the hormone binding domain. Many of these plasmid constructs were engineered to contain a polyhistidine tag on the N-terminus, which was used in the purification of the proteins.

Expression tests of four of the new constructs are shown in Figure 2. This figure shows that some protein is being made by these constructs under control of the inducible promoter (evident by comparing the "Uninduced" and "Induced" lanes for each construct.) A Western blot of these gels (not shown) with a polyclonal antibody specific to the C-terminal 30Kd of the Human Estrogen Receptor was used to confirm the expression of the ER constructs, and identify the position of these expression products (shown as white arrows in the Figure). However, these proteins are insoluble and remain in the pellet fraction

after ultracentrifugation. The "DE" and "DEF" constructs showed a moderate level of protein expression, but the protein likewise remained in the insoluble fraction during purification. The "CDE" and "CDEF" constructs failed to express at a significant level when initially transformed into BL21 or BL21-pLysS lines. Western blots using an antibody to the C-terminal portion of the ER showed that a small amount of protein was produced. Two sequential arginine codons rarely used in *E.coli* were noticed in the "C" domain, which may have been responsible for the expression failure. A helper plasmid containing an arginyl-tRNA (Brinkmann et al., 1989) was co-transfected with the constructs, resulting in more efficient expression through this region. Expression of the "CDE" and "CDEF" constructs did not, however, yield significant quantities of protein. The "ABCDE" construct also expressed protein with the helper plasmid present, but it was rapidly degraded, and isolation of the unclipped product was not successful. The full-length ER construct did not express any observable protein in this system.

Because we could not express soluble protein in the *E. coli* expression system, we chose to solubilize the different protein products in 8M urea, followed by refolding. Purification of the solubilized material was attempted by binding the polyhistidine tagged proteins to a Ni-NTA column. The "CDEF" and "DEF" protein constructs aggregated on the column despite the presence of 7M urea, and elution with imidazole yielded only microgram amounts of protein, which were insufficient for a crystallographic study. Attempts to refold the protein on the column by slowly removing the urea prior to the elution with imidazole failed to change the tendency of the protein to aggregate. Refolding of the "DE" and "CDE" constructs was not attempted. A summary of these results is shown in Table 1.

During the tenth month of the first year we learned that several other groups had succeeded in crystallizing the HBD from homologous receptors and that the structure determinations were essentially complete. Some of these results have recently been published (Bourquet et al., 1995; Wagner et al., 1995; Renaud et al., 1995). Moreover, another group has obtained crystals of the ER-HBD itself (P.Sigler, personal communication). These structures demonstrate that the last series of constructs we made should contain the entire domains of interest. Since we failed to make soluble protein with these constructs in a *E.coli* expression system, the next logical step for this project would be to start over

using a cell culture based expression system. We therefore decided to switch the emphasis of the project in order not to compromise Dr. Nolte's training in X-ray crystallography as applied to cancer related problems since starting up a new expression system for the estrogen receptor constructs would take a significant period of time.

2.TFIIIA: The transcription factor IIIA (TFIIIA) from Xenopus laevis binds both DNA and RNA. The discovery of multiple zinc finger motifs in TFIIIA pointed to a complex mode of protein-DNA interaction. Each of these motifs is an independently folded protein domain that recognizes a specific DNA sequence of three base pairs. Nine tandem zinc fingers are present in TFIIIA, allowing the protein to bind to an extensive region of the 5S RNA gene promoter.

The first six zinc fingers of TFIIIA have been expressed in E. coli and reconstituted with a duplex of two 31 base-pair synthetic DNA strands whose sequence derives from the Xenopus oocyte 5S RNA gene promoter. Crystals of the complex were grown using "native" DNA and one of two different substituted DNAs, in which three thymidine nucleotides were modified to contain iodine at specific locations. Several diffraction data sets have been collected from native and derivative crystals with the data in some cases extending to beyond 3.2Å resolution (Table 2). The work is at the stage of interpretation of difference-Patterson functions. Crystals for additional data collection have been grown.

Many genes with key roles in cancer are regulated by zinc-finger containing proteins. Mutations within the recently identified BRCA1 gene (Miki et al.,1994) have been correlated to susceptibility of breast cancer in individuals. This gene contains a zinc finger-like domain at it amino terminus. Understanding the principles of DNA recognition by members of this family will be important for establishing a picture of their transcriptional control. In particular, the feasibility of efforts to find drugs that target gene-specific transcriptional regulation will depend on such understanding.

Src-family kinases. We have determined the structures of the regulatory domains of the lymphocyte tyrosine kinase, Lck (Eck et al., 1993; Eck et al., 1994) We are currently engaged in efforts to crystallize the auto-inhibited state of the full-length, multi-domain molecule. Src-like kinases of this sort are important in a variety of tumors, and they appear to modify proteins that have a role in

breast cancer (see, for example, Arnold et al., 1995). Upon completion of the TFIIIA project, Dr. Nolte will build upon the work on Lck kinases by carrying out parallel studies of the related Fyn structure or by attempting expression and crystallization of ZAP-70, the immediate downstream target of Lck. These studies would not commence until year 3 of the project.

### Conclusion:

The revised SOW requested here for the coming year should see the completion of a crystallographic study of a complex between (TFIIIA fingers 1-6) and its cognate DNA. This structure will yield significant new information about modes of specific DNA recognition by this family of transcription factors. Other work in our laboratory on gene regulation at complex promoter/enhancers will then be poised to capitalize as the results. Dr. Nolte will himself complete this period of training and prepare to launch independent directions, by turning his attention to the cancer-related Src-like tyrosine kinases and to structural analysis of corresponding signal transduction pathways. Parallel efforts in our laboratory are proving successful at expression and crystallization of domains of the Lck kinase, and we will expand this effort to include a close homologue Fyn, as well as the downstream "effector" of Lck, ZAP-70.

## **References**

- Arnold, S.F., Obourn, J.D., Jaffe, H., and Notides, A.C. "Phosphorylation of the human estrogen receptor on tyrosine 537 in vivo and by Src family kinases in vitro". *Molecular Endocrinology* 9, 24-33 (1995).
- Bourguet, W., Ruff, M., Chambon, P., Gronmeyer, H., and Moras, D. "Crystal Structure of the ligand-binding domain of the human nuclear receptor RXR $\alpha$ ". *Nature* 375, 377-382 (1995).
- Brinkmann, U., Mattes, R.E., and Buckel, P. "High level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product". *Gene* 85, 109-114 (1989).
- Eck, M.J., Atwell, S.K., Shoelson, S.E. and Harrison, S.C. "Structure of the regulatory domains of the Src-family tyrosine kinase p56lck". *Nature*, 368, (1994).
- Eck, M.J., Shoelson, S.E., and Harrison, S.C. "Recognition of a high affinity phosphotyrosyl peptide by the Src homology domain of p56lck". *Nature*, 362, (1993).
- Harrison, S.C. "A structural taxonomy of DNA-binding domains". *Nature*, 353, (1991).
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S. et al. "A strong candidate for the breast and ovarian cancer susceptibility gene". *Science* 266, 66-71 (1994).
- Renaud, J.P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. "Crystal Structure of the RAR- $\gamma$  ligand-binding domain bound to all-trans retinoic acid". *Nature* 278, 681-689 (1995).
- Steitz, T.A., "Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding". *Quarterly Reviews of Biophysics*, 23, 205-280 (1990).
- Wagner, R.L., Apriletti, J.W., McGrath, M.E., West, B.E., Baxter, J.D., and Fletterick, R.J. "A structural role for hormone in the thyroid hormone receptor". *Nature* 378:690-696 (1995).

Table 1 : Summary of Estrogen Receptor Expression Results

Construct	Molecular Weight	Protein Expression	Protein Soluble	Protein Refolding
ABCDEF	65 Kd	No	---	---
ABCDE	61 Kd	Rapidly Degraded	No	---
CDE	41 Kd	Poor	No	?
CDEF	46 Kd	Poor	No	No
DE	34 Kd	Moderate	No	?
DEF	38 Kd	Moderate	No	No

Table 2 : Summary of TF3a X-Ray Data

Crystal	Unit Cell (Å and degrees)	Resolution Range (Å)	R-sym	X-ray Source
"Native"	63.8x64.6x78.0 90.1, 92.1, 102.6	20.0 - 4.0	8.8	Rotating Anode
Derivative 1	64.3x64.4x78.3 90.0, 93.0, 102.5	20.0 - 3.5	8.2	Rotating Anode
Derivative 2	63.3x63.5x76.5 90.1, 93.5, 102.6	20.0 - 3.1	7.3	NSLS - X12B

R-sym: conventional crystallographic agreement factor for symmetry-related intensities

NSLS-X12B: beam line X12B at National Synchrotron Light Source, Brookhaven National Laboratory



Figure 1: Summary of Estrogen Receptor Expression Constructs

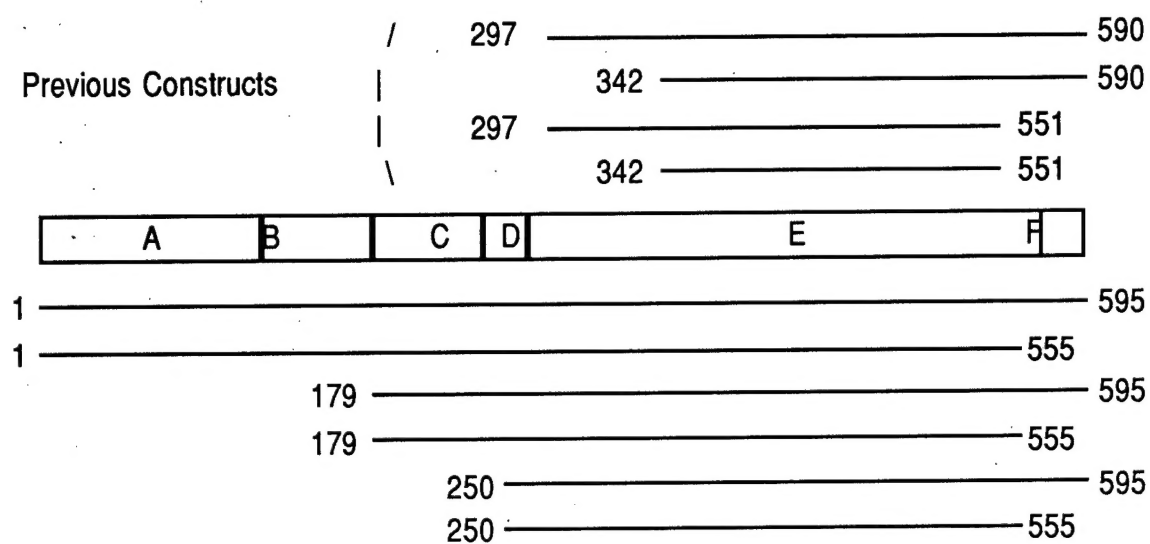


Figure 2 : Expression and Solubility Tests of Estrogen Receptor Constructs

This figure shows the expression of four of the new Estrogen receptor constructs. For each of the four constructs the following is shown. 1) Uninduced : Transfected *E. coli* without the addition of IPTG. 2) Induced : Transfected *E. coli* four hours after IPTG addition. 3) Whole Cell : Sonicated *E. coli* after four hour induction. 4) Soluble : Soluble portion of induced *E. coli* lysate (by Ultracentrifugation). 5) Insoluble : Pellet from Ultracentrifugation of induced *E. coli* Lysate.

